

Aprendizagem 2023

**Lab 3: Bayesian learning**

**Practical exercises**

1. **Probability theory**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A | B | C | D |
| 1 | 1 | 1 | 0 | 0 |
| 2 | 1 | 1 | 1 | 0 |
| 3 | 0 | 0 | 0 | 1 |
| 4 | 0 | 0 | 0 | 1 |
| 5 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 |

1. Consider the following registry where an experiment is repeated six times and four events (A, B, C and D) are detected.

Considering frequentist estimates, compute:

1. Considering the following two-dimensional measurements {(-2,2),(-1,3),(0,1),(-2,1)}.
2. What are the maximum likelihood parameters of a multivariate Gaussian distribution for this set of points?
3. A picture containing shape

   Description automatically generatedWhat is the shape of the Gaussian?

Draw it approximately using a contour map.

1. **Bayesian learning**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | y**1** | y**2** | y**3** | y**4** | y**5** | class |
| 1 | 1 | 1 | 0 | 1 | 0 | 1 |
| 2 | 1 | 1 | 1 | 0 | 0 | 0 |
| 3 | 0 | 1 | 1 | 1 | 0 | 0 |
| 4 | 0 | 0 | 0 | 1 | 1 | 0 |
| 5 | 1 | 0 | 1 | 1 | 1 | 1 |
| 6 | 0 | 0 | 1 | 0 | 0 | 1 |
| 7 | 0 | 0 | 0 | 0 | 1 | 1 |

1. Consider the following dataset where:

* 0: False and 1: True
* y1: Fast processing
* y2: Decent Battery
* y3: Good Camera
* y4: Good Look and Feel
* y5: Easiness of Use
* class: iPhone

And the query vector

1. Using Bayes’ rule, without making any assumptions, compute the posterior probabilities for the query vector. How is it classified?

According to our estimated likelihoods, the denominators are equal to zero. Posteriors are not defined and, thus, we cannot classify the input. A small training sample is not enough to decide under a classic Bayes rule.

1. What is the problem of working without assumptions?

Insufficient data to construct a meaningful joint distribution, e.g. applicable for datasets with high dimensionality or low size (small sample).

1. Compute the class for the same query vector under the naive Bayes assumption.

Label (not an iPhone).

1. Consider the presence of missings. Under the same naive Bayes assumption, how do you classify

Label .

|  |  |  |  |
| --- | --- | --- | --- |
|  | weight (kg) | height (cm) | NBA player |
| 1 | 170 | 160 | 0 |
| 2 | 80 | 220 | 1 |
| 3 | 90 | 200 | 1 |
| 4 | 60 | 160 | 0 |
| 5 | 50 | 150 | 0 |
| 6 | 70 | 190 | 1 |

1. Consider the following dataset

And the query vector

1. Compute the most probable class for the query vector assuming that the likelihoods are 2-dimensional Gaussians.

Classified as an NBA player.

1. Compute the most probable class for the query vector, under the Naive Bayes assumption, using 1-dimensional Gaussians to model the likelihoods

Classified as an NBA player.

1. Assuming training examples with *m* features and a binary class.
2. How many parameters do you have to estimate considering features are Boolean and:
3. no assumptions about how the data is distributed
4. naive Bayes assumption

One parameter for the prior

Considering the classic Bayesian model: we need parameters to estimate , hence .

Considering the naïve Bayes: we need to estimate . Since there are 2 classes and features, we have parameters for the likelihoods. The total number of parameters is .

1. How many parameters do you have to estimate considering features are numeric and:
2. multivariate Gaussian assumption
3. naive Bayes with Gaussian assumption

Similarly, one parameter for the prior, *.*

A multivariate Gaussian to estimate the likelihood requires a mean vector and a covariance matrix. For variables, the mean vector has parameters. The covariance is a matrix. However, the matrix is symmetric so, we only need to count the diagonal and upper diagonal part of the matrix, i.e. . In this context, the total number of parameters is .

Considering the naïve Bayes: we need to estimate , requiring the fitting of a (univariate) Gaussian distribution with two parameters: and . Since there are 2 classes and *m* features, we have parameters for the likelihoods. The total number of parameters is .

**Programming quests**

Resources: *Classification* and *Evaluation* notebooks available at the course’s webpage

1. Reuse the **sklearn** code from last lab where we learnt a decision tree in the *breast.w* data:
   1. apply the naïve Bayes classifier with default parameters
   2. compare the accuracy of both classifiers using a 10-fold cross-validation
2. Consider the accuracy estimates collected under a 5-fold CV for two predictive models M1 and M2, *acc*M1=(0.7,0.5,0.55,0.55,0.6) and *acc*M2=(0.75,0.6,0.6,0.65,0.55).

Using **scipy** (<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_rel.html>), assess whether the differences in predictive accuracy are statistically significant.